MODERATE DEVIATIONS IN A CLASS OF STABLE BUT NEARLY UNSTABLE PROCESSES

FRÉDÉRIC PROÏA

Abstract. We consider a stable but nearly unstable autoregressive process of any order. The bridge between stability and instability is expressed by a time-varying companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \to 1$. In that framework, we establish a moderate deviation principle for the empirical covariance only relying on the elements of A_n through $1 - \rho(A_n)$ and, as a by-product, we establish a moderate deviation principle for the OLS estimator when Γ, the renormalized asymptotic variance of the process, is invertible. Finally, when Γ is singular, we also provide a compromise in the form of a moderate deviation principle for a penalized version of the estimator. Our proofs essentially rely on troncations and m–dependent sequences with unbounded m.

1. Introduction and Assumptions

Unit root issues have long been crucial in time series econometrics and have therefore focused a great deal of research studies. This sudden demarcation between stability and instability is responsible for many inference problems in linear time series (see Brockwell and Davis [4] for a detailed overview of the linear stochastic processes). The remarkable works of Chan and Wei [7] encompass, in a much more general context, the now well-known fact that the least squares estimator is $\sqrt{n}$–consistent with Gaussian behavior when the underlying autoregressive process is stable, whereas it is n–consistent with asymmetrical distribution when the process is unstable. This rather abrupt change in the rate of convergence and in the asymptotic distribution certainly motivated the wide range of unit root testing procedures, but it also paved the way for studies based on time-varying coefficients. In a nearly unstable autoregressive process, we do not focus on a parameter θ satisfying $|\theta| < 1$ or $|\theta| = 1$ but, instead, the parameter is considered as a sequence $|\theta_n| < 1$ such that $|\theta_n| \to 1$. This sample size dependent structure allows a continuity between stability and instability. For example, Phillips and Magdalinos [20] treat the case where the coefficient is in a $O(\kappa_n^{-1})$ neighborhood of the unit root with $\kappa_n = o(n)$ and, amongst other results, prove a central limit theorem for the estimator at the rate $\sqrt{n\,\kappa_n}$, illustrating thereby the bridge existing between the stable rate $\sqrt{n}$ and the unstable rate n, stemming from the time-invariant framework. In the same vein, let us also mention the work of Chan and Wei [6], or natural generalizations like the study of Phillips and Lee [19] related to vector autoregressions and the recent unified theory of Buchmann and Chan [5], focused on the general nearly unstable autoregressive processes. Our paper is precisely based on the latter topic, in a sense that will be precised in good time.

Given a parametric generating process, the precision of the estimation is usually assessed by its rate of convergence and the deviations can be seen as a natural continuation after a central limit theorem or even a law of iterated logarithm. Roughly speaking, they may be

Key words and phrases. Nearly unstable autoregressive process, Moderate deviation principle, OLS estimation, Asymptotic behavior, Unit root.

used to estimate the exponential decline of the probability of tail events such as any non-zero distance between the estimator and the parameter of interest. We refer the reader to Dembo and Zeitouni [8] to access all the mathematical formalization. Since the 1980s, numerous authors have worked on large and/or moderate deviations in a time series context under many and varied hypotheses. Without claiming to be exhaustive, one can mention Donsker and Varadhan [10] and Bercu *et al.* [2] on stationary Gaussian processes and quadratic forms, Worms [21] on Markov chains and regression models, Bercu [1] on first-order Gaussian stable, unstable and explosive processes, Mas and Menneteau [15] on Hilbertian processes, Djellout *et al.* [9] on non-linear functionals of moving average processes, Wu and Zhao [22] on stationary non-linear processes, Miao and Shen [16] on general autoregressive processes or, more recently, Bitseki Penda *et al.* [3] on first-order processes with correlated errors, and, of course, all the references inside.

In that paper, we investigate the moderate deviations of the estimate in stable but nearly unstable autoregressions. This can be seen as a full generalization of the recent work of Miao, Wang and Yang [17], focused on the univariate case. More precisely, for a fixed $n \geqslant 1$, let the process given for some $p \geqslant 1$ and $k \in \{1, \ldots, n\}$ by

$$X_{n,\,k} = \sum_{i=1}^{p} \theta_{n,\,i}\, X_{n,\,k-i} + \varepsilon_k$$

where $(\varepsilon_k)_k$ is a sequence of zero-mean i.i.d. random variables. In an equivalent way, we can consider the vector expression

$$\Phi_{n,\,k} = A_n\, \Phi_{n,\,k-1} + E_k \tag{1.1}$$

where $E_k = (\varepsilon_k, 0, \ldots, 0)^T$ is a p–vectorial noise, $\Phi_{n,\,k} = (X_{n,\,k}, \ldots, X_{n,\,k-p+1})^T$ and

$$A_n = \begin{pmatrix} \theta_{n,\,1} & \theta_{n,\,2} & \ldots & \theta_{n,\,p} \\ & I_{p-1} & & 0 \end{pmatrix} \tag{1.2}$$

is the $p \times p$ companion matrix of the autoregressive process. If $(E_k)_k$ has a finite variance, it is well-known that $(\Phi_{k,\,n})_k$ is a second-order stationary process having the causal form

$$\Phi_{n,\,k} = \sum_{\ell=0}^{+\infty} A_n^{\ell}\, E_{k-\ell} \tag{1.3}$$

when $\rho(A_n) < 1$, that is, when the largest modulus of its eigenvalues is less than 1 (see *e.g.* Thm. 11.3.1 of [4] and the fact that each eigenvalue of A_n is the inverse of a zero of the autoregressive polynomial of the process). Since $(\varepsilon_k)_k$ is an i.i.d. sequence, the process is strictly stationary with mean zero and variance given by

$$\Gamma_n = \sigma^2 \sum_{\ell=0}^{+\infty} A_n^{\ell}\, K_p\, (A_n^T)^{\ell} \tag{1.4}$$

where, for convenience, we will denote in the whole study

$$K_p = \begin{pmatrix} 1 & 0 \\ 0 & 0_{p-1} \end{pmatrix} \quad \text{and} \quad U_p = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{1.5}$$

the $p \times p$ matrix with 1 at the top left and 0 elsewhere, and its first column standing for the first vector of the canonical basis of $\mathbb{R}^p$. As a consequence of the causal expression above, the initial vector $\Phi_{n,0}$ is not arbitrary and has to share the distribution of the process. This also implies the relation

$$\Gamma_n = A_n \, \Gamma_n \, A_n^T + \sigma^2 \, K_p. \tag{1.6}$$

As will be largely developped throughout the study, Γ_n is finite for all $n \geqslant 1$ but, as n increases, $|||\Gamma_n||| \to +\infty$. The keystone matrix Γ obtained after a correct standardization of Γ_n is the renormalized asymptotic variance of the process. Before we start, we define a matrix that will also prove to be crucial to our results,

$$B_n = I_{p^2} - A_n \otimes A_n. \tag{1.7}$$

We are now going to introduce and comment the hypotheses that will be needed, though not always simultaneously, in the whole paper. Section 2 is devoted to our main results : two theorems related to moderate deviations, a set of explicit examples and some additional comments and conclusions. Finally, in Section 3 divided into numerous subsections, we will prove all our results, step by step.

Remark. We denote by $\|\cdot\|$ the Euclidean vector norm and by $|||\cdot|||$ the spectral matrix norm. Other norms may be used, in which case an appropriated subscript is added. Moreover, we will always denote by $\langle \cdot, \cdot \rangle$ the usual inner product of the Euclidean space $\mathbb{R}^d$ for any $d \geqslant 1$.

1.1. **Hypotheses.** First of all, we present the hypotheses that we retain.

(H_1) *Gaussian integrability condition.* There exists $\alpha > 0$ such that

$$\mathbb{E}\Big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\Big] < +\infty$$

where ε_1 represents the zero-mean i.i.d. sequence $(\varepsilon_k)_k$ of variance $\sigma^2 > 0$ and fourth-order moment $\tau^4 > 0$.

(H_2) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

with distinct eigenvalues $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_1| = \rho(A)$, and the top right element of A is non-zero.

(H_3) *Spectral radius of the companion matrix.* For all $n \geqslant 1$, $\rho(A_n) < 1$. In addition,

$$\lim_{n \to +\infty} \rho(A_n) = \rho(A) = 1.$$

(H_4) *Renormalization.* We have the convergences

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_*} = H \quad \text{and} \quad \lim_{n \to +\infty} (1 - \rho(A_n))\, |||B_n^{-1}|||_* = h$$

for some matrix norm, where H is a $p^2 \times p^2$ non-zero matrix and $h > 0$.

(H_5) *Moderate deviations.* The moderate deviations scale (b_n) satisfies

$$\lim_{n \to +\infty} b_n = +\infty \quad \text{and} \quad \lim_{n \to +\infty} \frac{\sqrt{n}\,(1 - \rho(A_n))^{\frac{3}{2}+\eta}}{b_n} = +\infty$$

for a small $\eta > 0$.

1.2. **Comments on the hypotheses.** First, in view of the link between the spectrum of the companion matrix and the zeros of the autoregressive polynomial, 0 cannot be an eigenvalue of A. Conceding in (H$_2$) that the limiting matrix has distinct eigenvalues is a matter of simplication of the reasonings. Indeed, A_n turns out to be diagonalizable for a sufficiently large n, and, as a companion matrix, it is well-known that the change of basis is done *via* a Vandermonde matrix having numerous nice properties (more details are given in Section 3.1, and a discussion on the case of multiple eigenvalues is provided in Section 2.3). The top right element of A_n is $\theta_{n,p}$, so assuming in (H$_2$) that $\theta_{n,p} \nrightarrow 0$ ensures that the limit process is still of order p. Moreover, note that, in (H$_4$), the invertibility of B_n for all n is guaranteed by (H$_3$). Indeed, $\rho(A_n \otimes A_n) = \rho^2(A_n) < 1$ (see *e.g.* Lem. 5.6.10 and Cor. 5.6.16 of [13]). In addition, we obviously have, for all $\ell \geqslant 0$,

$$\rho(A_n^\ell) = \rho^\ell(A_n) \leqslant |||A_n^\ell|||$$

so that we get

$$\frac{1}{1-\rho(A_n)} \leqslant \sum_{\ell=0}^{+\infty} |||A_n^\ell||| = L_n \tag{1.8}$$

giving a lower bound for L_n. Similarly,

$$\frac{1}{(1-\rho(A_n))^2} \leqslant \sum_{\ell=0}^{+\infty} (\ell+1)\, |||A_n^\ell||| = M_n. \tag{1.9}$$

However, an exact upper bound for these sums may be difficult to reach and may require stringent conditions on the elements of A_n. We refer the reader to Lemma 3.1 where, under (H$_2$) and (H$_3$), some asymptotic upper bounds are established. We also refer to Section 2.2 where the explicit calculations in terms of some examples shall help to understand the rates involved in the hypotheses. Now for a fixed $n \geqslant 1$, let

$$\mu_n = \rho(A_n) + \frac{1-\rho(A_n)}{2} = \frac{\rho(A_n)+1}{2}.$$

Clearly, $\rho(A_n) < \mu_n < 1$. Hence, according to Prop. 2.3.15 of [11], for all $n \geqslant 0$, there exists a constant $c_n > 0$ such that, for all $\ell \geqslant 0$, $|||A_n^\ell||| \leqslant c_n\, \mu_n^\ell$ so that

$$L_n \leqslant \frac{c_n}{1-\mu_n} < +\infty \quad \text{and} \quad M_n \leqslant \frac{c_n}{(1-\mu_n)^2} < +\infty.$$

Letting n tend to infinity, it follows from (H$_3$) and (H$_4$) that

$$\lim_{n\to+\infty} |||B_n^{-1}||| = \lim_{n\to+\infty} L_n = \lim_{n\to+\infty} M_n = +\infty. \tag{1.10}$$

Finally, it will be established in good time that there is a limiting matrix Γ such that

$$\lim_{n\to+\infty} \frac{\Gamma_n}{|||B_n^{-1}|||_*} = \Gamma \tag{1.11}$$

where $|||\cdot|||_*$ is the matrix norm of (H$_4$).

Remark. To facilitate the reading, we consider from now on that the matrix norm $|||\cdot|||_*$ is identified in (H$_4$), and we will only note $|||\cdot|||$ in the sequel.

2. Main results

This section contains two theorems that constitute the main results of the paper. The first of them is quite long to establish and will need numerous technical lemmas, but the second one will essentially be deduced as a corollary of the first one. Subsequently, we provide some explicit examples for a better understanding and an easier interpretation of the hypotheses together with some graphics showing the evolution of the processes and the estimation of the autoregressive parameter. At the end of the section, we discuss the case of multiple eigenvalues. But before anything else, let us recall the definition of the large and moderate deviation principles (see Sec. 1.2 of [8] for more details). In the sequel, a speed is considered as a positive sequence increasing to infinity.

Definition. A sequence of random variables $(U_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a large deviation principle (LDP) with speed (a_n) and rate I if there is a lower semicontinuous mapping $I : \mathcal{X} \to \bar{\mathbb{R}}^+$ such that :

- for any closed set $F \in \mathcal{B}$,
$$\limsup_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in F) \leqslant - \inf_{x \in F} I(x),$$
- for any open set $G \in \mathcal{B}$,
$$- \inf_{x \in G} I(x) \leqslant \liminf_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in G).$$

In particular, if the infimum of I coincides on the interior H° and the closure $\bar{H}$ of some $H \in \mathcal{B}$, then
$$\lim_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in H) = - \inf_{x \in H} I(x).$$

Definition. A sequence of random variables $(V_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a moderate deviation principle (MDP) with speed (b_n^2) and rate I if there is a speed (v_n) with $\frac{v_n}{b_n} \to +\infty$ such that $(\frac{v_n}{b_n} V_n)_n$ satisfies a large deviation principle of speed (b_n^2) and rate I.

2.1. **Moderate deviations.** We now consider an observable trajectory $X_{n,-p+1}, \ldots, X_{n,n}$ for some fixed $n \geqslant 1$, and use it to provide an estimation of the parameter. It is well-known that the ordinary least squares (OLS) estimator of $\theta_n = (\theta_{n,1}, \ldots, \theta_{n,p})^T$ is given by

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1} X_{n,k} \quad \text{where} \quad S_{n-1} = \sum_{k=1}^{n} \Phi_{n,k-1} \Phi_{n,k-1}^T. \tag{2.1}$$

The first result is dedicated to the empirical variance $\frac{S_n}{n}$.

Theorem 2.1. *Under hypotheses (H$_1$)–(H$_5$), the sequence*
$$\left(\frac{\sqrt{n}\,(1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \operatorname{vec}\left(\frac{1}{n} \sum_{k=1}^{n} (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) \right) \right)_{n \geqslant 1}$$
satisfies an LDP with speed (b_n^2) and a rate function $I_\Gamma : \mathbb{R}^{p^2} \to \bar{\mathbb{R}}^+$ defined as
$$I_\Gamma(x) = \begin{cases} \frac{1}{2h^3} \langle x, \Upsilon^\dagger x \rangle & \text{for } x \in \operatorname{Im}(\Upsilon) = \operatorname{Im}(\Upsilon^\dagger) \\ +\infty & \text{otherwise} \end{cases}$$
where $\Upsilon^\dagger$ is the generalized inverse of Υ explicitely given in (3.18)*, and h comes from (H$_4$).*

Proof. See Section 3.2.5. ☐

Remark. Through vectorization, this MDP is established on $\mathbb{R}^{p^2}$ in order to avoid any confusion in the notations, but we might work in $\mathbb{R}^{p\times p}$ as well. The associated rate function would only require a slight modification of the proof.

Remark. To be punctilious, we may add a small $\epsilon > 0$ to the diagonal of S_{n-1} to ensure that it is non-sigular for all $n \geqslant 1$ without disturbing the asymptotic behavior.

When the variance Γ given in (1.11) is invertible, we establish the MDP for the OLS in the theorem that follows. However, when it is not the case, there are some technical complications and, to reach an intermediate result, we need to introduce a penalized version of the OLS. For a small $\pi \geqslant 0$, define

$$\widehat{\theta}_n^{\,\pi} = (S_{n-1}^{\pi})^{-1} \sum_{k=1}^{n} \Phi_{n,\,k-1}\, X_{n,\,k} \quad \text{where} \quad S_{n-1}^{\pi} = S_{n-1} + \pi\, n\, |||B_n^{-1}|||\, I_p \tag{2.2}$$

with possibly $\pi = 0$ if Γ is invertible, in which case it is clearly the standard OLS given above, but necessarily $\pi > 0$ otherwise. Consider also the penalized version of the variance and the corrected parameter

$$\Gamma_\pi = \Gamma + \pi\, I_p \quad \text{and} \quad \theta_n^{\pi} = (S_{n-1}^{\pi})^{-1} S_{n-1}\, \theta_n. \tag{2.3}$$

By construction, Γ is, at worst, non-negative definite and for $\pi > 0$, Γ_π turns out to be invertible. The same goes for S_{n-1}^{π}.

Theorem 2.2. *Under hypotheses* (H_1)*–*(H_5)*, for all* $\pi > 0$*, the sequence*

$$\left(\frac{\sqrt{n}}{b_n\, (1-\rho(A_n))^{\frac{1}{2}}} \left(\widehat{\theta}_n^{\,\pi} - \theta_n^{\pi} \right) \right)_{n\,\geqslant\,1}$$

satisfies an LDP with speed (b_n^2) *and a rate function* $I_\theta^{\pi} : \mathbb{R}^p \to \bar{\mathbb{R}}^+$ *defined as*

$$I_\theta^{\pi}(x) = \begin{cases} \frac{h}{2\,\sigma^2}\, \langle x, \Gamma_\pi\, \Gamma^{\dagger}\, \Gamma_\pi\, x \rangle & \text{for } x \in \mathrm{Im}(\Gamma_\pi^{-1}\,\Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

where $\Gamma^{\dagger}$ *is the generalized inverse of the variance* Γ *given in* (1.11)*,* Γ_π *is the penalized variance given in* (2.3) *and* h *comes from* (H_4)*, respectively. If in addition* Γ *is invertible, then the sequence*

$$\left(\frac{\sqrt{n}}{b_n\, (1-\rho(A_n))^{\frac{1}{2}}} \left(\widehat{\theta}_n - \theta_n \right) \right)_{n\,\geqslant\,1}$$

satisfies an LDP with speed (b_n^2) *and a rate function* $I_\theta : \mathbb{R}^p \to \mathbb{R}^+$ *defined as*

$$I_\theta(x) = \frac{h}{2\,\sigma^2}\, \langle x, \Gamma\, x \rangle.$$

Proof. See Section 3.2.6. ☐

To sum up, this result shows that, when Γ is invertible, the OLS satisfies an MDP, and even when Γ is singular, one may reach as a compromise an MDP for a penalized estimator close to the OLS. In the same vein, notice also that

$$\lim_{\pi\,\to\,0^+} I_\theta^{\pi}(x) = I_\theta(x).$$

Remark. In the stable case where $\rho(A_n) = \rho(A) < 1$, we simply have $(1-\rho(A_n))\,|||B_n^{-1}||| = h$ and $\Gamma_n\,|||B_n^{-1}|||^{-1} = \Gamma$ for all $n \geqslant 1$. By contraction, the MDP of Theorem 2.2 coincides with the one of Thm. 3 of [21] when Γ is invertible.

2.2. **Some explicit examples.** Before giving some examples, we can already note that (H_5) implies $\sqrt{n}\,(1-\rho(A_n)) \to +\infty$. Thus, necessarily, the convergence $1-\rho(A_n) \to 0$ cannot occur with an exponential rate, this is the reason why we focus on polynomial rates of the form $1-\rho(A_n) = c\,n^{-\alpha}$ for some $c > 0$ in this section. Accordingly, in all the examples below, (H_5) is only possible when $0 < \alpha < \frac{1}{3+2\eta} < \frac{1}{3}$. Thus, one cannot expect a sequence of coefficients moving too fast toward instability. The domain of validify of the speed of the MDP will be

$$1 \ll b_n \ll n^{\frac{1-(3+2\eta)\alpha}{2}} \ll \sqrt{n}.$$

2.2.1. *Univariate case with one nearly unit root.* Suppose that $p = 1$. Then, (H_2) and (H_3) imply that $|\theta_n| < 1$ and $\theta_n \to \pm 1$. We also have $B_n = 1-\theta_n^2$ and (H_4) can be expressed like

$$\lim_{n\to+\infty} \frac{B_n^{-1}}{|B_n^{-1}|} = 1 \quad \text{and} \quad \lim_{n\to+\infty} (1-|\theta_n|)\,|B_n^{-1}| = \frac{1}{2}.$$

A straightforward calculation shows that

$$\Gamma_n = \frac{\sigma^2}{1-\theta_n^2} \quad \text{and} \quad \Gamma = \sigma^2 > 0$$

so that we can choose $\pi = 0$. The standard cases – illustrated on Figure 1 – are $\theta_n = 1 - c_1\,n^{-\alpha}$ for the positive unit root and $\theta_n = -1 + c_2\,n^{-\alpha}$ for the negative unit root, with $c_1, c_2 > 0$ and $\alpha > 0$. The rate function associated with Theorem 2.2 is $I_\theta(x) = \frac{x^2}{4}$, which corresponds to Prop. 2.1 of [17]. Indeed, their rate $x \mapsto \frac{x^2}{2}$ is associated to an LDP with the renormalization $(1-\theta_n^2)^{\frac{1}{2}}$ whereas our normalization is $(1-|\theta_n|)^{\frac{1}{2}}$. By contraction, the asymptotic factor $\sqrt{2}$ explains the difference.

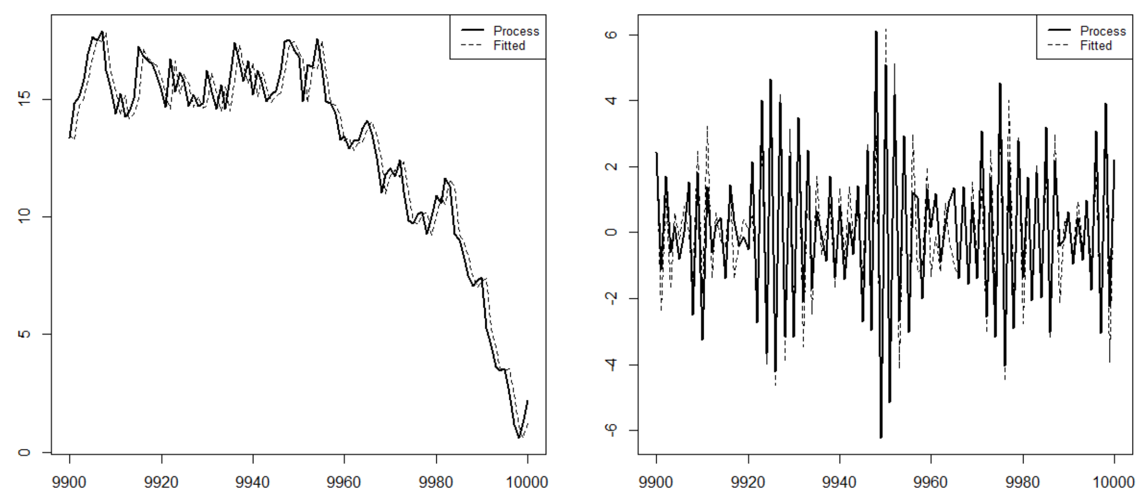

FIGURE 1. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0,1)$ innovations. The setting is $c_1 = 0.1$ and $\alpha = 0.32$ on the left, $c_2 = 0.1$ and $\alpha = 0.32$ on the right.

2.2.2. *Bivariate case with one nearly unit root.* Suppose now that $p = 2$ and $\mathrm{sp}(A) = \{\pm 1, \lambda\}$ with $|\lambda| < 1$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} \lambda + 1 - c\, n^{-\alpha} & -\lambda\, (1 - c\, n^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

whose eigenvalues are $1 - c\, n^{-\alpha}$ and λ – this is illustrated on Figure 2. For $c > 0$ and $\alpha > 0$, (H_2) and (H_3) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{2\, c\, (\lambda - 1)^2} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} \big(n^{\alpha} \; + O(1) \big)$$

whence we obtain

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_1} = \frac{1}{4} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} \qquad \text{and} \qquad \lim_{n \to +\infty} n^{-\alpha}\, |||B_n^{-1}|||_1 = \frac{2}{c\, (\lambda - 1)^2}$$

so (H_4) is satisfied with the 1–norm. The choice $\pi = 0$ is impossible, and we finally find

$$\Gamma_{\pi} = \frac{\sigma^2}{4} \begin{pmatrix} 1 + \frac{4}{\sigma^2}\, \pi & 1 \\ 1 & 1 + \frac{4}{\sigma^2}\, \pi \end{pmatrix}.$$

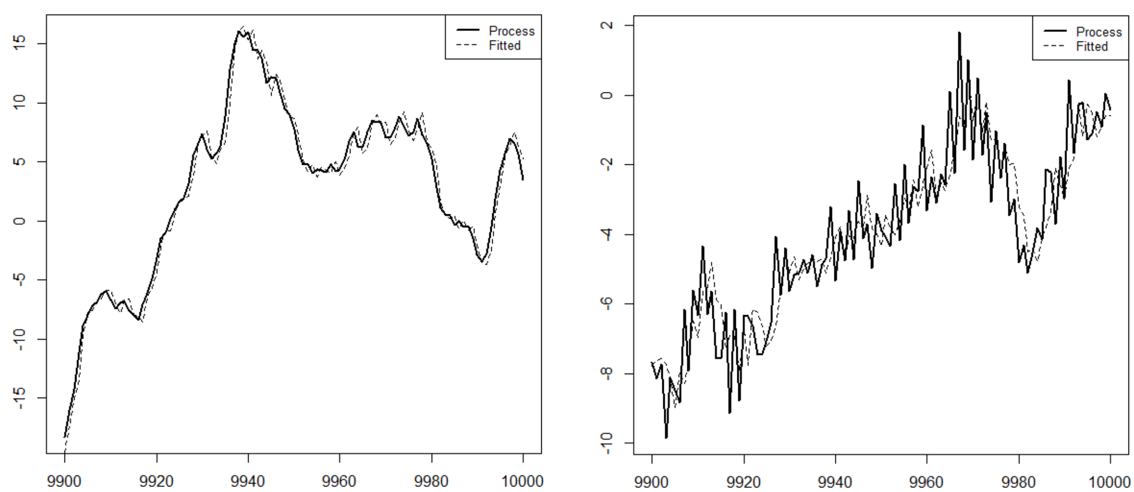

FIGURE 2. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 10^{-5}$ and $\mathcal{N}(0, 1)$ innovations. The setting is $\lambda = 0.5$, $c = 0.1$ and $\alpha = 0.32$ on the left, $\lambda = -0.67$, $c = 0.2$ and $\alpha = 0.25$ on the right.

2.2.3. *Bivariate case with two nearly unit roots.* Following the same lines, suppose that $p = 2$ and $\mathrm{sp}(A) = \{-1, 1\}$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} (c_2 - c_1) n^{-\alpha} & (1 - c_1\, n^{-\alpha})(1 - c_2\, n^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

whose eigenvalues are $1 - c_1\, n^{-\alpha}$ and $-1 + c_2\, n^{-\alpha}$ – this is illustrated on Figure 3. For $c_1, c_2 > 0$ and $\alpha > 0$, (H$_2$) and (H$_3$) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{8\, c_1\, c_2} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix} \big(n^{\alpha} \ + O(1)\big)$$

whence we obtain

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_1} = \frac{1}{2\,(c_1 + c_2) + 2\,|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

Moreover,

$$\lim_{n \to +\infty} n^{-\alpha}\, |||B_n^{-1}|||_1 = \frac{(c_1 + c_2 + |c_2 - c_1|)}{4\, c_1\, c_2}$$

so (H$_4$) is satisfied with the 1–norm. The choice $\pi = 0$ is possible and we finally find

$$\Gamma = \frac{\sigma^2}{2\,(c_1 + c_2) + 2\,|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

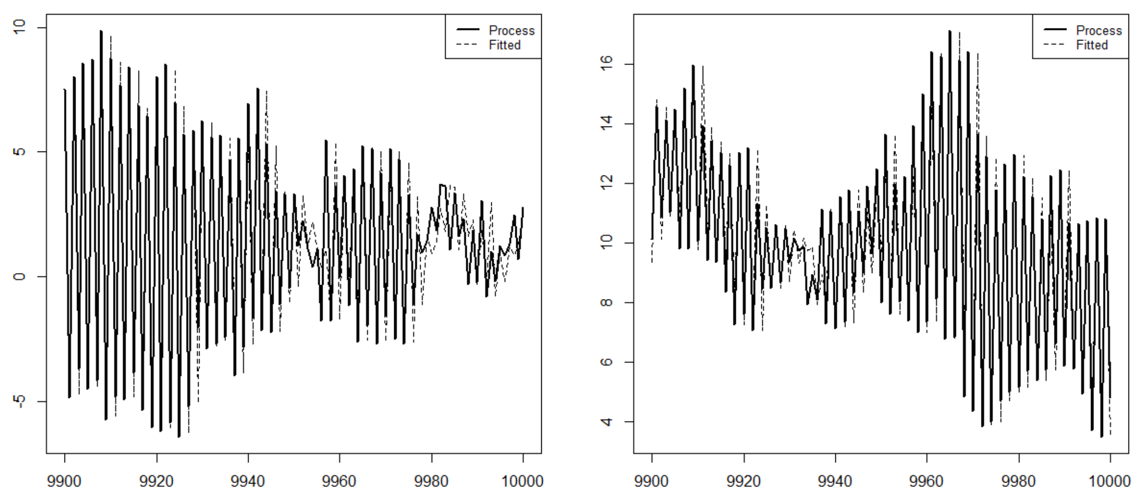

FIGURE 3. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0, 1)$ innovations. The setting is $c_1 = 0.1$, $c_2 = 0.2$ and $\alpha = 0.32$ on the left, $c_1 = 0.01$, $c_2 = 0.01$ and $\alpha = 0.25$ on the right.

2.3. **Discussion on multiple eigenvalues and conclusion.** The distinct eigenvalues assumption (H$_2$) is clearly sufficient, since A_n turns out to be diagonalizable for the large values of n and since, as we will see in the proof of Lemma 3.1, the change of basis matrices P_n and P_n^{-1} can be uniformly bounded, most of our intermediate results stemming from that fact. A less stringent formulation of (H$_2$) could be :

(H$'_2$) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

and the top right element of A is non-zero. In addition, there exists a rank n_0 such that, for all $n > n_0$, A_n is diagonalizable and the change of basis matrix P_n satisfies $|||P_n||| \leqslant C_{st}$ and $|||P_n^{-1}||| \leqslant C_{st}$.

In general, multiple eigenvalues may not falsify our reasonings, except when the multiplicity concerns the eigenvalues whose modulus tends to 1. Indeed, the coefficients of $|||A_n^{\ell}|||$ may grow faster in that case. Consider the simple bivariate example where

$$A_n^{\ell} \;=\; \begin{pmatrix} a_{11,\,\ell} & a_{12,\,\ell} \\ a_{21,\,\ell} & a_{22,\,\ell} \end{pmatrix} \;=\; \begin{pmatrix} \theta_{n,\,1}\, a_{11,\,\ell-1} + \theta_{n,\,2}\, a_{11,\,\ell-2} & \theta_{n,\,2}\, a_{11,\,\ell-1} \\ \theta_{n,\,1}\, a_{21,\,\ell-1} + \theta_{n,\,2}\, a_{21,\,\ell-2} & \theta_{n,\,2}\, a_{21,\,\ell-1} \end{pmatrix}.$$

Then, it is not hard to solve this linear difference equation whose characteristic roots are the eigenvalues of A_n. In case of multiplicity, the top left term takes the form of

$$a_{11,\,\ell} = (c_n + d_n\, \ell)\, \rho^{\ell}(A_n)$$

and even if $|c_n| \leqslant C_{st}$ and $|d_n| \leqslant C_{st}$ for n large enough, it follows that

$$\sum_{\ell=0}^{+\infty} |||A_n^{\ell}||| \sim \frac{C_{st}}{(1-\rho(A_n))^2}.$$

That invalidates all our reasonings and, in that case, new approaches are needed to potentially reach the moderate deviations. From our viewpoint, this is the main weakness of the paper. As it is already observed in [7], multiple unit roots located at 1 change the rate of convergence of the OLS, we conjecture that the same phenomenon occurs here and that a larger power should come with $1 - \rho(A_n)$ in the renormalization.

To sum up, this study is a wide generalization of [17] and, although not complete in virtue of the latter remark, it covers most of the MDP issues for the estimation in the stable but nearly unstable case. Large deviations are undoubtedly a very useful and challenging study to carry out, naturally extending this one. However, to the best of our knowledge, it is not even entirely treated in the stable case $\rho(A_n) = \rho(A) < 1$, clearly revealing the complexity of the problem. It could also be complicated but stimulating to investigate the exponential, and not only polynomial, neighborhood of the unit root.

3. Technical proofs

In all the proofs, C_{st} denotes a generic positive constant that is not necessarily identical from one line to another. We will frequently use the fact that $\|\mathrm{vec}(\cdot)\| = |||\cdot|||_F \leqslant C_{st}\, |||\cdot|||$. For asymptotic equivalences, $f_n \asymp g_n$ means that both $f_n = O(g_n)$ and $g_n = O(f_n)$ whereas $f_n \sim g_n$ stands for $\frac{f_n}{g_n} \to 1$.

3.1. **Some linear algebra tools.** Thereafter, we denote by $\lambda_1, \ldots, \lambda_p$ the (distinct) eigenvalues of A and $\lambda_{n,\,1}, \ldots, \lambda_{n,\,p}$ those of A_n, in descending order of modulus. We start by establishing two lemmas that will prove to be very useful in the sequel.

Lemma 3.1. *Under hypotheses* (H_2) *and* (H_3)*, as* n *tends to infinity,*

$$\sum_{\ell=0}^{+\infty} |||A_n^{\ell}||| \asymp \frac{1}{1-\rho(A_n)} \quad \text{and} \quad \sum_{\ell=0}^{+\infty} (\ell+1)\, |||A_n^{\ell}||| \asymp \frac{1}{(1-\rho(A_n))^2}.$$

Proof. The lower bounds are established in Section 1.2. For the upper bounds, fix

$$\delta = \frac{2}{|\lambda_p|}, \qquad \epsilon_1 = \frac{1}{2} \min_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left| \frac{1}{\lambda_i} - \frac{1}{\lambda_j} \right| \quad \text{and} \quad \epsilon_2 = 2 \max_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left| \frac{1}{\lambda_i} - \frac{1}{\lambda_j} \right|.$$

According to Thm. 2.4.9.2 of [13], (H_2) implies the existence of a rank $n_0 = n_0(\delta, \epsilon_1, \epsilon_2)$ such that, for all $n > n_0$, the eigenvalues of A_n satisfy

$$0 < \max_{1 \leqslant i \leqslant p} \left| \frac{1}{\lambda_{n,\,i}} \right| < \delta \tag{3.1}$$

and

$$\epsilon_1 < \min_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left| \frac{1}{\lambda_{n,\,i}} - \frac{1}{\lambda_{n,\,j}} \right| < \max_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left| \frac{1}{\lambda_{n,\,i}} - \frac{1}{\lambda_{n,\,j}} \right| < \epsilon_2. \tag{3.2}$$

Let P_n be a change of basis matrix in the diagonalization of A_n. Then, since A_n is a companion matrix, a standard choice would be

$$P_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{\lambda_{n,\,1}} & \frac{1}{\lambda_{n,\,2}} & \cdots & \frac{1}{\lambda_{n,\,p}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{n,\,1}^{p-1}} & \frac{1}{\lambda_{n,\,2}^{p-1}} & \cdots & \frac{1}{\lambda_{n,\,p}^{p-1}} \end{pmatrix}. \tag{3.3}$$

This Vandermonde matrix is invertible if and only if $\lambda_{n,\,i} \neq \lambda_{n,\,j}$ for all $i \neq j$ (see *e.g.* Sec. 0.9.11 of [13]). In that case, P_n^{-1} is closely related to the Lagrange interpolating polynomials given, for $i \in \{1, \ldots, p\}$, by

$$L_i(X) = \frac{\prod_{j \neq i} (X - \frac{1}{\lambda_{n,\,j}})}{\prod_{j \neq i} (\frac{1}{\lambda_{n,\,i}} - \frac{1}{\lambda_{n,\,j}})}.$$

Precisely, the i–th row of P_n^{-1} contains the coefficients of $L_i(X)$ in the basis $(1, X, \ldots, X^{p-1})$ of $\mathbb{R}_{p-1}[X]$, *i.e.*

$$P_n^{-1} = \left(\frac{p_{n,\,i,\,j}}{\prod_{j \neq i} (\frac{1}{\lambda_{n,\,i}} - \frac{1}{\lambda_{n,\,j}})} \right)_{1 \leqslant i,j \leqslant p} \tag{3.4}$$

where the relation $\prod_{j \neq i} (X - \frac{1}{\lambda_{n,\,j}}) = p_{n,\,i,\,1} + p_{n,\,i,\,2}\, X + \ldots + p_{n,\,i,\,p}\, X^{p-1}$ enables to identify each $p_{n,\,i,\,j}$. Combining (3.1) and (3.2), it follows that, for all $n > n_0$,

$$|||P_n|||_1 \,\leqslant\, p\, (1 + \delta + \ldots + \delta^{\,p-1}) \,\leqslant\, C_{st}.$$

We also have $|||P_n^{-1}|||_1 \leqslant C_{st}$ since $\epsilon_1^{p-1} < \prod_{j \neq i} |\frac{1}{\lambda_{n,\,i}} - \frac{1}{\lambda_{n,\,j}}| < \epsilon_2^{p-1}$ and since $p_{n,\,i,\,j}$ is a finite combination of sums and products of $\frac{1}{\lambda_{n,\,1}}, \ldots, \frac{1}{\lambda_{n,\,p}}$. To sum up, for all $\ell \geqslant 0$ and $n > n_0$,

$$A_n^{\ell} = P_n\, D_n^{\,\ell}\, P_n^{-1} \quad \text{where} \quad D_n = \mathrm{diag}(\lambda_{n,\,1}, \ldots, \lambda_{n,\,p}).$$

Consequently,

$$
\begin{aligned}
|||A_n^\ell||| &= |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + |||P_n\, D_n^\ell\, P_n^{-1}|||\, \mathbb{1}_{\{n > n_0\}} \\
&\leqslant |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + |||P_n|||\, |||P_n^{-1}|||\, |||D_n^\ell|||\, \mathbb{1}_{\{n > n_0\}} \\
&\leqslant |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, \rho^\ell(A_n)\, \mathbb{1}_{\{n > n_0\}}. \qquad (3.5)
\end{aligned}
$$

It only remains to sum over ℓ and to let n tend to infinity to reach the first result. Similarly,

$$
(\ell+1)\, |||A_n^\ell||| \leqslant (\ell+1)\, |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, (\ell+1)\, \rho^\ell(A_n)\, \mathbb{1}_{\{n > n_0\}}
$$

so that we get the second result following the same lines. □

Lemma 3.2. *Under hypotheses (H_2) and (H_3), we have the convergence*

$$
\lim_{n \to +\infty} A_n^{w_n} = 0
$$

for any rate w_n satisfying $w_n\,(1-\rho(A_n)) \to +\infty$.

Proof. Consider the rank n_0 introduced in the proof of Lemma 3.1. Then, according to the inequality (3.5),

$$
|||A_n^{w_n}||| \leqslant |||A_n^{w_n}|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, \rho^{w_n}(A_n)\, \mathbb{1}_{\{n > n_0\}} \qquad (3.6)
$$

where the invertible and uniformly bounded matrices P_n and P_n^{-1} are given in (3.3) and (3.4), respectively. We also have

$$
\lim_{n \to +\infty} \rho^{w_n}(A_n) = \lim_{n \to +\infty} \mathrm{e}^{-w_n\,(1-\rho(A_n))} = 0 \qquad (3.7)
$$

from the hypothesis on w_n. It remains to let n tend to infinity in the above inequality. □

3.2. **Proofs of the main results.** To direct the proof of our theorems, first of all it is convenient to express the empirical variance of the process as

$$
\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}(\Phi_{n,k}\,\Phi_{n,k}^T - \Gamma_n) &= \frac{1}{n}\sum_{k=1}^{n} A_n\,\Phi_{n,k-1}\,\Phi_{n,k-1}^T\,A_n^T + \frac{1}{n}\sum_{k=1}^{n} A_n\,\Phi_{n,k-1}\,E_k^T \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} E_k\,\Phi_{n,k-1}^T\,A_n^T + \frac{1}{n}\sum_{k=1}^{n} E_k\,E_k^T - \Gamma_n \\
&= \frac{1}{n}\sum_{k=1}^{n} \Delta_{n,k} + \frac{1}{n}\sum_{k=1}^{n} A_n\,(\Phi_{n,k}\,\Phi_{n,k}^T - \Gamma_n)\,A_n^T - \frac{T_n}{n}
\end{aligned}
$$

where the variance Γ_n is given in (1.4),

$$
\Delta_n = \frac{1}{n}\sum_{k=1}^{n}(A_n\,\Phi_{n,k-1}\,E_k^T + E_k\,\Phi_{n,k-1}^T\,A_n^T + E_k\,E_k^T + A_n\,\Gamma_n\,A_n^T - \Gamma_n) \qquad (3.8)
$$

and the residual term is

$$
T_n = A_n\,(\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T)\,A_n^T.
$$

Then, solving this generalized Sylvester equation (Lem. 2.1 of [14]), considering the invertibility of B_n in (1.7) which is justified at the beggining of Section 1.2, we reach the decomposition

$$\mathrm{vec}\left(\frac{1}{n}\sum_{k=1}^{n}(\Phi_{n,k}\,\Phi_{n,k}^T-\Gamma_n)\right)=B_n^{-1}\,\mathrm{vec}(\Delta_n)-\frac{B_n^{-1}\,\mathrm{vec}(T_n)}{n}. \tag{3.9}$$

Let us now reason step by step, *via* some intermediate results.

3.2.1. *Exponential moments of the squared initial value.* We recall that, from the causal form (1.3) of the process,

$$\Phi_{n,0}=\sum_{\ell=0}^{+\infty}A_n^{\ell}\,E_{-\ell}.$$

Lemma 3.3. *Under hypothesis* (H_1),

$$\mathbb{E}\left[\exp\left(\frac{\alpha}{L_n^2}\,|||\Phi_{n,0}\,\Phi_{n,0}^T|||\right)\right]<+\infty$$

where L_n *is given in* (1.8).

Proof. By Cauchy-Schwarz inequality,

$$\begin{aligned}|||\Phi_{n,0}\,\Phi_{n,0}^T||| \;\leqslant\; \|\Phi_{n,0}\|^2 \;&\leqslant\; \left(\sum_{\ell=0}^{+\infty}\|A_n^{\ell}\,E_{-\ell}\|\right)^2\\ &\leqslant\; \left(\sum_{\ell=0}^{+\infty}|||A_n^{\ell}|||^{\frac{1}{2}}\,|||A_n^{\ell}|||^{\frac{1}{2}}\,\|E_{-\ell}\|\right)^2 \;\leqslant\; L_n\sum_{\ell=0}^{+\infty}|||A_n^{\ell}|||\,\varepsilon_{-\ell}^2.\end{aligned}$$

Moreover, from Jensen's inequality, for all $\lambda>0$,

$$\exp\left(\frac{\lambda}{L_n}\sum_{\ell=0}^{+\infty}|||A_n^{\ell}|||\,\varepsilon_{-\ell}^2\right)\;\leqslant\;\frac{1}{L_n}\sum_{\ell=0}^{+\infty}|||A_n^{\ell}|||\,\mathrm{e}^{\lambda\,\varepsilon_{-\ell}^2}$$

using $\frac{|||A_n^0|||}{L_n}+\frac{|||A_n^1|||}{L_n}+\ldots=1$. Taking the expectation and choosing $\lambda=\alpha$ given in (H_1), we deduce that

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{\alpha}{L_n^2}\,|||\Phi_{n,0}\,\Phi_{n,0}^T|||\right)\right]\;&\leqslant\;\frac{1}{L_n}\sum_{\ell=0}^{+\infty}|||A_n^{\ell}|||\,\mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_{-\ell}^2}\big]\\ &=\;\mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]<+\infty.\end{aligned} \tag{3.10}$$

□

3.2.2. *Exponential convergence of the residual term.* The residual term in the decomposition (3.9) is given by

$$R_n=\frac{B_n^{-1}\,\mathrm{vec}(A_n\,(\Phi_{n,n}\,\Phi_{n,n}^T-\Phi_{n,0}\,\Phi_{n,0}^T)\,A_n^T)}{n}. \tag{3.11}$$

Lemma 3.4. *Under hypotheses* (H_1)*–*(H_5)*, for all* $r>0$,

$$\lim_{n\to+\infty}\frac{1}{b_n^2}\ln\mathbb{P}\left(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,\|R_n\|\geqslant r\right)=-\infty.$$

Proof. First, note that

$$\begin{aligned}
\|R_n\| &\leqslant \frac{|||B_n^{-1}|||\,\|\mathrm{vec}(A_n\,(\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T)\,A_n^T)\|}{n} \\
&\leqslant \frac{C_{st}\,|||B_n^{-1}|||\,|||A_n|||^2\,|||\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T|||}{n} \\
&\leqslant \frac{C_{st}\,|||B_n^{-1}|||\,|||A_n|||^2\,(|||\Phi_{n,n}\,\Phi_{n,n}^T||| + |||\Phi_{n,0}\,\Phi_{n,0}^T|||)}{n}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{P}\bigg(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,\|R_n\| \geqslant r\bigg) &= \mathbb{P}\bigg(\|R_n\| \geqslant \frac{r\,b_n\,(1-\rho(A_n))^{-\frac{3}{2}}}{\sqrt{n}}\bigg) \\
&\leqslant 2\,\mathbb{P}\bigg(|||\Phi_{n,0}\,\Phi_{n,0}^T||| \geqslant \frac{r\,b_n\sqrt{n}\,(1-\rho(A_n))^{-\frac{3}{2}}}{2\,C_{st}\,|||A_n|||^2\,|||B_n^{-1}|||}\bigg) \\
&\leqslant 2\,\mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]\exp\bigg(-\frac{r\,\alpha\,b_n\sqrt{n}\,(1-\rho(A_n))^{-\frac{3}{2}}}{2\,C_{st}\,|||A_n|||^2\,|||B_n^{-1}|||\,L_n^2}\bigg)
\end{aligned}$$

where L_n is given in (1.8), using Markov's inequality, the reasoning in the proof of Lemma 3.3 and the fact that, from the strict stationarity of the process, $\Phi_{n,0}\,\Phi_{n,0}^T$ and $\Phi_{n,n}\,\Phi_{n,n}^T$ share the same distribution. Hence, for a sufficiently large n,

$$\begin{aligned}
\frac{1}{b_n^2}\ln\mathbb{P}\bigg(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,\|R_n\| \geqslant r\bigg) &\leqslant \frac{\ln 2 + \ln\mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]}{b_n^2} - \frac{r\,\alpha\,\sqrt{n}\,(1-\rho(A_n))^{-\frac{3}{2}}}{2\,C_{st}\,b_n\,|||A_n|||^2\,|||B_n^{-1}|||\,L_n^2} \\
&\leqslant \frac{\ln 2 + \ln\mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]}{b_n^2} - C_{st}\,\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}
\end{aligned}$$

since $|||B_n^{-1}|||^{\frac{1}{2}} \sim \sqrt{h}\,(1-\rho(A_n))^{-\frac{1}{2}}$ from (H$_4$), $L_n^2 = O((1-\rho(A_n))^{-2})$ from Lemma 3.1 and since, from (H$_2$), $|||A_n|||$ converges. Finally, letting n tend to infinity, (H$_1$) and (H$_5$) conclude the proof. $\square$

3.2.3. *The truncated sequence.* In all the sequel, we define the rate

$$m_n = \left\lfloor \left(\frac{1}{1-\rho(A_n)}\right)^{\frac{3+3\eta}{3+2\eta}} \right\rfloor \tag{3.12}$$

and we note from (H$_3$)–(H$_5$) that

$$\lim_{n\to+\infty} m_n(1-\rho(A_n)) = +\infty \quad \text{and} \quad \lim_{n\to+\infty} \frac{b_n\,|||B_n^{-1}|||^{\frac{1}{2}}\,m_n^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0. \tag{3.13}$$

Following the idea of [17], we are going to use m_n as a truncation parameter. Consider

$$\Psi_{n,k} = \sum_{\ell=0}^{m_n-2} A_n^{\ell}\,E_{k-\ell} \tag{3.14}$$

as an approximation of $\Phi_{n,k}$ in its causal form (1.3). We also define the truncated version of the summands of $\Delta_{n,k}$ in (3.8) as

$$\zeta_{n,k} = A_n\,\Psi_{n,k-1}\,E_k^T + E_k\,\Psi_{n,k-1}^T\,A_n^T + E_k\,E_k^T + A_n\,\Gamma_n\,A_n^T - \Gamma_n. \tag{3.15}$$

The process $(B_n^{-1}\,\mathrm{vec}(\zeta_{n,k}))_k$ is strictly stationary and m_n–dependent, according to Def. 6.4.3 of [4]. Let us study some properties of this process.

Lemma 3.5. *Under hypotheses (H$_1$)–(H$_4$), we can find a constant $c_\alpha > 0$ such that, for a sufficiently large n,*

$$\mathbb{E}\bigg[\exp\Big(c_\alpha\,|||B_n^{-1}|||^{-1}\sum_{\ell=0}^{w_n}|||A_n^\ell\,E_{-\ell}\,E_1^T|||\Big)\bigg] \,\leqslant\, \mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]$$

for any rate w_n satisfying $w_n\,(1-\rho(A_n)) \to +\infty$.

Proof. By Hölder's inequality,

$$\mathbb{E}\bigg[\exp\Big(c_\alpha\,|||B_n^{-1}|||^{-1}\sum_{\ell=0}^{w_n}|||A_n^\ell\,E_{-\ell}\,E_1^T|||\Big)\bigg] \,\leqslant\, \mathbb{E}\bigg[\exp\Big(c_\alpha\,|||B_n^{-1}|||^{-1}\sum_{\ell=0}^{w_n}|||A_n^\ell|||\,\varepsilon_1^2\Big)\bigg].$$

Moreover, for the rank n_0 and the uniformly bounded matrices P_n and P_n^{-1} introduced in the proof of Lemma 3.1,

$$\begin{aligned}\sum_{\ell=0}^{w_n}|||A_n^\ell||| &= \sum_{\ell=0}^{n_0}|||A_n^\ell||| + \sum_{\ell=n_0+1}^{w_n}|||A_n^\ell|||\\ &= \sum_{\ell=0}^{n_0}|||A_n^\ell||| + \sum_{\ell=n_0+1}^{w_n}|||P_n\,D_n^\ell\,P_n^{-1}||| \,\leqslant\, C_{st}\left(1+\frac{1-\rho^{\,w_n}(A_n)}{1-\rho(A_n)}\right)\end{aligned}$$

as soon as $w_n > n_0$. Thus,

$$|||B_n^{-1}|||^{-1}\sum_{\ell=0}^{w_n}|||A_n^\ell||| \,\leqslant\, C_{st}\left(|||B_n^{-1}|||^{-1}+\frac{1-\rho^{\,w_n}(A_n)}{|||B_n^{-1}|||\,(1-\rho(A_n))}\right).$$

Finally, (H$_4$), (1.10) and (3.7) lead, for the large values of n, to

$$|||B_n^{-1}|||^{-1}\sum_{\ell=0}^{w_n}|||A_n^\ell||| \,\leqslant\, C_{st}.$$

It remains to choose $c_\alpha = \frac{\alpha}{C_{st}}$. □

Lemma 3.6. *Under hypotheses (H$_2$)–(H$_4$), for all $n \geqslant 1$ and $k \in \{1,\ldots,n\}$,*

$$\mathbb{E}[\mathrm{vec}(\zeta_{n,k})] = 0 \quad \textit{and} \quad \mathbb{C}\mathrm{ov}(\mathrm{vec}(\zeta_{n,k}),\mathrm{vec}(\zeta_{n,j})) = \begin{cases} 0 & \textit{for } k \neq j\\ \Upsilon_n & \textit{for } k = j\end{cases}$$

where the $p^2 \times p^2$ covariance Υ_n can be explicitely built in terms of σ^2, A_n and B_n. In addition,

$$\lim_{n\,\to\,+\infty}\frac{B_n^{-1}\,\Upsilon_n\,(B_n^{-1})^T}{|||B_n^{-1}|||^{\,3}} = \Upsilon$$

where the non-zero limiting matrix Υ is given in (3.18).

Proof. We will use in the sequel K_p and U_p defined in (1.5). Let $\mathcal{F}_k = \sigma(\varepsilon_\ell,\,\ell \leqslant k)$ be the σ–algebra of the events occurring up to time k. Then, it is easy to see that

$$\begin{aligned}\mathbb{E}[\mathrm{vec}(\zeta_{n,k})] &= \mathbb{E}[\,\mathbb{E}[\mathrm{vec}(\zeta_{n,k})\,|\,\mathcal{F}_{k-1}]\,]\\ &= \sigma^2\,\mathrm{vec}(K_p)+\mathrm{vec}(A_n\,\Gamma_n\,A_n^T-\Gamma_n) \,=\, 0\end{aligned}$$

in virtue of (1.6). For $k > j$, by direct calculation,

$$\begin{aligned} \mathbb{E}[\text{vec}(\zeta_{n,k})\,\text{vec}^T(\zeta_{n,j})] &= \mathbb{E}[\,\mathbb{E}[\text{vec}(\zeta_{n,k})\,\text{vec}^T(\zeta_{n,j})\,|\,\mathcal{F}_{k-1}]\,] \\ &= \mathbb{E}[\,(\mathbb{E}[\text{vec}(A_n\,\Psi_{n,k-1}\,E_k^T) + \text{vec}(E_k\,\Psi_{n,k-1}^T\,A_n^T)|\,\mathcal{F}_{k-1}] \\ &\qquad + \sigma^2\,\text{vec}(K_p) + \text{vec}(A_n\,\Gamma_n\,A_n^T - \Gamma_n))\,\text{vec}^T(\zeta_{n,j})] \;=\; 0 \end{aligned}$$

and the same is true for $j > k$ since $(\mathbb{E}[\text{vec}(\zeta_{n,k})\,\text{vec}^T(\zeta_{n,j})])^T = \mathbb{E}[\text{vec}(\zeta_{n,j})\,\text{vec}^T(\zeta_{n,k})] = 0$. Now for $k = j$, a tedious but straightforward calculations leads to

$$\begin{aligned} \mathbb{E}[\text{vec}(\zeta_{n,k})\,\text{vec}^T(\zeta_{n,k})] &= \sigma^2\,K_p \otimes (A_n\,\mathbb{E}[\Psi_{n,k-1}\Psi_{n,k-1}^T]\,A_n^T) \\ &\quad + \sigma^2\,U_p \otimes (A_n\,\mathbb{E}[\Psi_{n,k-1}\Psi_{n,k-1}^T]\,A_n^T) \otimes U_p^T \\ &\quad + \sigma^2\,U_p^T \otimes (A_n\,\mathbb{E}[\Psi_{n,k-1}\Psi_{n,k-1}^T]\,A_n^T) \otimes U_p \\ &\quad + \sigma^2\,(A_n\,\mathbb{E}[\Psi_{n,k-1}\Psi_{n,k-1}^T]\,A_n^T) \otimes K_p \\ &\quad + (\tau^4 - \sigma^4)\,\text{vec}(K_p)\,\text{vec}^T(K_p) \;=\; \Upsilon_n. \end{aligned} \tag{3.16}$$

To give an explicit expression of Υ_n, it suffices to observe that the truncated expression (3.14) has a variance given by

$$\Gamma_{n,m_n} = \mathbb{E}[\Psi_{n,k-1}\Psi_{n,k-1}^T] = \sigma^2 \sum_{\ell=0}^{m_n-2} A_n^\ell\,K_p\,(A_n^T)^\ell$$

so that

$$\begin{aligned} \text{vec}(\Gamma_{n,m_n}) &= \sigma^2 \sum_{\ell=0}^{m_n-2} (A_n \otimes A_n)^\ell\,\text{vec}(K_p) \\ &= \sigma^2\,B_n^{-1}\,(I_{p^2} - (A_n \otimes A_n)^{m_n-1})\,\text{vec}(K_p). \end{aligned}$$

Let us now look at the asymptotic behavior of Υ_n correctly renormalized. First, we have the convergence

$$\lim_{n\,\to\,+\infty} (A_n \otimes A_n)^{m_n-1} = 0$$

coming from the identity $(A_n \otimes A_n)^{m_n-1} = A_n^{m_n-1} \otimes A_n^{m_n-1}$ and Lemma 3.2. Together with (H_4), this implies

$$\lim_{n\,\to\,+\infty} \frac{\text{vec}(\Gamma_{n,m_n})}{|||B_n^{-1}|||} = \sigma^2\,H\,\text{vec}(K_p).$$

In the end of the proof, we call vec^{-1} the vectorization inverse operator (namely, in our context, the reconstruction of a $p \times p$ matrix from its vectorization of size p^2). Then,

$$\lim_{n\,\to\,+\infty} \frac{\Gamma_{n,m_n}}{|||B_n^{-1}|||} = \sigma^2\,\text{vec}^{-1}(H\,\text{vec}(K_p)) = \Gamma. \tag{3.17}$$

Combining (3.16) with (3.17) and (H_4),

$$\Upsilon = \sigma^2\,H\,(K_p \otimes \Gamma^A + U_p \otimes \Gamma^A \otimes U_p^T + U_p^T \otimes \Gamma^A \otimes U_p + \Gamma^A \otimes K_p)\,H^T \tag{3.18}$$

where, for convenience, we denote $\Gamma^A = A\,\Gamma\,A^T$. $\square$

Remark. As a by-product, we also obtain, following the same lines,

$$\lim_{n \to +\infty} \frac{\Gamma_n}{|||B_n^{-1}|||} = \Gamma$$

where Γ_n is given in (1.4), which justifies (1.11). The variance $\Gamma_{n,\,m_n}$ defined above may be seen as the truncated version of Γ_n.

3.2.4. *The remainder of the truncation.* We denote by

$$\Lambda_n = \frac{1}{n} \sum_{k=1}^{n} (A_n \, (\Phi_{n,\,k-1} - \Psi_{n,\,k-1}) \, E_k^T + E_k \, (\Phi_{n,\,k-1} - \Psi_{n,\,k-1})^T \, A_n^T) \tag{3.19}$$

the remainder of the truncation of Δ_n in (3.8) made *via* (3.15). Our objective is now to establish the following lemma.

Lemma 3.7. *Under hypotheses (H_1)–(H_5), for all $r > 0$,*

$$\lim_{n \to +\infty} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\frac{\sqrt{n}\, (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \, \| B_n^{-1} \, \mathrm{vec}(\Lambda_n) \| \geqslant r \bigg) = -\infty.$$

Proof. Clearly, both terms in the definition of (3.19) are similar and we will only work on the first one. From the causal expression (1.3) and the truncation (3.14), we note that

$$\begin{aligned} \sum_{k=1}^{n} A_n \, (\Phi_{n,\,k-1} - \Psi_{n,\,k-1}) \, E_k^T &= \sum_{k=1}^{n} \sum_{\ell=m_n-1}^{+\infty} A_n^{\ell+1} \, E_{k-1-\ell} \, E_k^T \\ &= A_n^{m_n} \sum_{\ell=0}^{+\infty} A_n^{\ell} \sum_{k=1}^{n} E_{k-\ell-m_n} \, E_k^T. \end{aligned}$$

Thus, with M_n given in (1.9) and applying Lem. 17 of [15] under (H_1),

$$\begin{aligned} &\mathbb{P}\bigg(\frac{1}{n} \bigg|\bigg|\bigg| \sum_{k=1}^{n} A_n \, (\Phi_{n,\,k-1} - \Psi_{n,\,k-1}) \, E_k^T \bigg|\bigg|\bigg| \geqslant r \, \frac{b_n}{\sqrt{n}} \, |||B_n^{-1}|||^{\frac{1}{2}} \bigg) \\ &\qquad \leqslant \ \mathbb{P}\bigg(\sum_{\ell=0}^{+\infty} |||A_n^{\ell}||| \, \bigg| \sum_{k=1}^{n} \varepsilon_{k-\ell-m_n} \, \varepsilon_k \bigg| \geqslant \sum_{\ell=0}^{+\infty} (\ell+1) \, |||A_n^{\ell}||| \, \frac{r \, b_n \sqrt{n} \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||} \bigg) \\ &\qquad \leqslant \ \sum_{\ell=0}^{+\infty} \mathbb{P}\bigg(\max_{1 \leqslant j \leqslant n} \bigg| \sum_{k=1}^{j} \varepsilon_{k-\ell-m_n} \, \varepsilon_k \bigg| \geqslant \frac{r \, (\ell+1) \, b_n \sqrt{n} \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||} \bigg) \\ &\qquad \leqslant \ C_{st} \sum_{\ell=0}^{+\infty} \exp\bigg(- \frac{r^2 \, b_n^2 \, n \, t_{n,\,\ell}^2}{\alpha_0 \, n + \beta_0 \, r \, t_{n,\,\ell} \, b_n \sqrt{n}} \bigg) \end{aligned} \tag{3.20}$$

for some $\alpha_0 > 0$ and $\beta_0 > 0$, where

$$t_{n,\,\ell} = \frac{(\ell+1) \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||}.$$

Our choice of m_n in (1.9), the properties of Lemma 3.1, (3.6) and our hypotheses on the rates of convergence lead, for n large enough, to

$$|||B_n^{-1}|||^{-\frac{1}{2}} \, M_n \, |||A_n^{m_n}||| \ \leqslant \ C_{st} \, (1 - \rho(A_n))^{-\frac{3}{2}} \, \rho^{m_n}(A_n) \ \longrightarrow \ 0$$

and obviously $t_{n,\ell} \to +\infty$. Hence, like in formula (3.11) of [17], there are some constants $\alpha_0' > 0$ and $\beta_0' > 0$ such that, for all $\ell \geqslant 0$ and the large values of n,

$$\begin{aligned} \frac{r^2\, n\, b_n^2\, t_{n,\ell}^2}{\alpha_0\, n + \beta_0\, r\, t_{n,\ell}\, b_n \sqrt{n}} & = \frac{r^2\, (\ell+1)\, b_n^2\, t_{n,\ell}}{\alpha_0\, |||B_n^{-1}|||^{-\frac{1}{2}}\, M_n\, |||A_n^{m_n}||| + r\, \beta_0\, (\ell+1)\, \frac{b_n}{\sqrt{n}}} \\ & \geqslant b_n^2\, t_{n,\ell}\, \frac{r^2}{\alpha_0' + r\, \beta_0'}. \end{aligned}$$

Going back to (3.20),

$$\begin{aligned} \sum_{\ell=0}^{+\infty} \exp\Big(- \frac{r^2\, b_n^2\, n\, t_{n,\ell}^2}{\alpha_0\, n + \beta_0\, r\, t_{n,\ell}\, b_n \sqrt{n}}\Big) & \leqslant \sum_{\ell=0}^{+\infty} \exp\Big(- b_n^2\, t_{n,\ell}\, \frac{r^2}{\alpha_0' + r\, \beta_0'}\Big) \\ & = \frac{\mathrm{e}^{-V_n}}{1 - \mathrm{e}^{-V_n}} \end{aligned}$$

where, for convenience, we note

$$V_n = \frac{r^2\, b_n^2\, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n\, |||A_n^{m_n}|||\, (\alpha_0' + r\, \beta_0')} \longrightarrow +\infty.$$

To sum up,

$$\begin{aligned} & \frac{1}{b_n^2} \ln \mathbb{P}\Bigg(\Big|\Big|\Big| \sum_{k=1}^{n} A_n\, (\Phi_{n,k-1} - \Psi_{n,k-1})\, E_k^T \Big|\Big|\Big| \geqslant r\, b_n \sqrt{n}\, |||B_n^{-1}|||^{\frac{1}{2}} \Bigg) \\ & \qquad \leqslant \frac{C_{st} - \ln(1 - \mathrm{e}^{-V_n})}{b_n^2} - \frac{V_n}{b_n^2} \\ & \qquad \leqslant \frac{C_{st} - \ln(1 - \mathrm{e}^{-V_n})}{b_n^2} - \frac{r^2\, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n\, |||A_n^{m_n}|||\, (\alpha_0' + r\, \beta_0')} \longrightarrow -\infty. \end{aligned}$$

This is clearly sufficient to achieve the proof since, from (H_4),

$$\begin{aligned} \frac{\sqrt{n}\, (1 - \rho(A_n))^{\frac{3}{2}}}{b_n}\, \| B_n^{-1}\, \mathrm{vec}(\Lambda_n) \| & \leqslant \frac{\sqrt{n}}{b_n}\, |||B_n^{-1}|||^{-\frac{1}{2}}\, \| \mathrm{vec}(\Lambda_n) \| \\ & \leqslant C_{st}\, \frac{\sqrt{n}}{b_n}\, |||B_n^{-1}|||^{-\frac{1}{2}}\, |||\Lambda_n||| \end{aligned}$$

for n large enough. □

3.2.5. *Proof of Theorem 2.1.* All the technical results of the previous sections are now going to be concretely used. Consider the sequence

$$\xi_{n,k} = \frac{B_n^{-1}\, \mathrm{vec}(\zeta_{n,k})}{|||B_n^{-1}|||^{\frac{3}{2}}} \tag{3.21}$$

where $\zeta_{n,k}$ is given in (3.15). The process $(\xi_{n,k})_k$ is also strictly stationary and m_n–dependent. Like in [18] or [17, suppl. mat.], let us extract an independent sequence from this process. For $j \in \{1, \ldots, j_n\}$, define

$$\xi_{n,j}' = \xi_{n,(j-1)m_n+1} + \ldots + \xi_{n,jm_n}$$

where $j_n = \lfloor \frac{n}{m_n} \rfloor$ and where m_n and its properties are given in (3.12). Then, $(\xi'_{n,\,j})_j$ is strictly stationary and 1–dependent. Next, for $t \in \{1, \ldots, t_n\}$, define

$$\xi''_{n,\,t} = \xi'_{n,\,(t-1)u_n+1} + \ldots + \xi'_{n,\,t u_n - 1}$$

where $t_n = \lfloor \frac{j_n}{u_n} \rfloor$ and u_n is another rate satisfying

$$\lim_{n \to +\infty} u_n = +\infty \quad \text{and} \quad \lim_{n \to +\infty} \frac{b_n\, |||B_n^{-1}|||^{\frac{1}{2}}\, (m_n\, u_n)^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0. \tag{3.22}$$

To be convinced that such a rate exists, one can exploit (3.13) and the fact that $|\ln f_n| \to +\infty$ and $f_n\, |\ln f_n|^a \to 0$ when $f_n \to 0$. The process $(\xi''_{n,\,t})_t$ is now i.i.d. and the rates satifies

$$\lim_{n \to +\infty} \frac{t_n\, u_n\, m_n}{n} = 1. \tag{3.23}$$

The reasoning of [17, suppl. mat.] does not suit us, so we need to reformulate the establishment of the MDP. First, by a Taylor-Lagrange expansion,

$$\exp\left(\left\langle \lambda,\, \frac{b_n}{\sqrt{n}}\, \xi''_{n,\,1} \right\rangle\right) = 1 + \frac{b_n}{\sqrt{n}}\, \langle \lambda,\, \xi''_{n,\,1} \rangle + \frac{b_n^2}{2\,n}\, \langle \lambda,\, \xi''_{n,\,1} \rangle^2 + \frac{b_n^3}{6\,n^{\frac{3}{2}}}\, \langle \lambda,\, \xi''_{n,\,1} \rangle^3\, \mathrm{e}^{\nu_n} \tag{3.24}$$

in which the remainder satisfies, for any $\alpha > 0$,

$$\begin{aligned} \mathrm{e}^{\alpha\,\nu_n} \ &< \ \exp\left(\frac{\alpha\, b_n}{\sqrt{n}}\, |\langle \lambda,\, \xi''_{n,\,1} \rangle|\right) \ \leqslant \ \exp\left(\frac{\alpha\, b_n}{\sqrt{n}}\, \|\lambda\| \sum_{\ell=1}^{m_n u_n} \|\xi_{n,\,\ell}\|\right) \\ &\leqslant \ \exp\left(C_{st}\, \frac{b_n}{\sqrt{n}}\, |||B_n^{-1}|||^{-\frac{1}{2}} \sum_{\ell=1}^{m_n u_n} |||\zeta_{n,\,\ell}|||\right). \end{aligned}$$

Now, the random variables $\|\xi_{n,\,\ell}\|$ sharing the same distribution for all $\ell \geqslant 0$, it follows from Hölder's inequality that,

$$\begin{aligned} \mathbb{E}\big[\mathrm{e}^{\alpha\,\nu_n}\big] \ &< \ \mathbb{E}\Big[\exp\left(C_{st}\, \frac{b_n\, m_n\, u_n}{\sqrt{n}}\, |||B_n^{-1}|||^{-\frac{1}{2}}\, |||\zeta_{n,\,1}|||\right)\Big] \\ &\leqslant \ \mathbb{E}\Big[\exp\left(C_{st}\, \frac{b_n\, |||B_n^{-1}|||^{\frac{1}{2}}\, m_n\, u_n}{\sqrt{n}}\, |||B_n^{-1}|||^{-1}\, |||\zeta_{n,\,1}|||\right)\Big] \ < \ +\infty \end{aligned} \tag{3.25}$$

for n large enough, using Lemma 3.5 with $m_n\,(1 - \rho(A_n)) \to +\infty$ stemming from (3.13), the convergence of $|||A_n|||$, (H$_1$) and treating all the terms of (3.15) similarly. Taking the expectation in (3.24) and exploiting the independence of the zero-mean process $(\xi''_{n,\,t})_t$, we obtain the decomposition

$$\begin{aligned} \frac{1}{b_n^2} \ln \mathbb{E}\Big[\exp\left(\left\langle \lambda,\, \frac{b_n}{\sqrt{n}} \sum_{\ell=1}^{n} \xi_{n,\,\ell} \right\rangle\right)\Big] \ &\sim \ \frac{t_n}{b_n^2}\, \ln \mathbb{E}\Big[\exp\left(\left\langle \lambda,\, \frac{b_n}{\sqrt{n}}\, \xi''_{n,\,1} \right\rangle\right)\Big] \\ &= \ \frac{t_n}{2\,n}\, \mathbb{E}\big[\langle \lambda,\, \xi''_{n,\,1} \rangle^2\big] + O\left(\frac{t_n\, b_n}{6\,n^{\frac{3}{2}}}\, \big|\mathbb{E}\big[\langle \lambda,\, \xi''_{n,\,1} \rangle^3\, \mathrm{e}^{\nu_n}\big]\big|\right) \end{aligned} \tag{3.26}$$

for we can see, as it is done in [18], that the residual term

$$\tau_n = \sum_{\ell=1}^{n} \xi_{n,\,\ell} - \sum_{\ell=1}^{t_n} \xi''_{n,\,\ell}$$

plays a negligible role in comparison to the main one. To eliminate the third-order term, we first look at the fourth-order moment of $\langle \lambda,\, \xi''_{n,1} \rangle$, that is

$$\mathbb{E}\big[\langle \lambda,\, \xi''_{n,1} \rangle^4\big] \ \leqslant \ \frac{C_{st}\, \|\lambda\|^4}{|||B_n^{-1}|||^2}\, \mathbb{E}\Big[\Big|\Big|\Big| \sum_{\ell=1}^{m_n u_n} \zeta_{n,\ell} \Big|\Big|\Big|^4\Big].$$

A long but standard calculation shows that

$$\begin{aligned}
\mathbb{E}\Big[\Big|\Big|\Big| A_n \sum_{\ell=1}^{m_n u_n} \Psi_{n,\ell-1}\, E_\ell^T \Big|\Big|\Big|^4\Big] \ &\leqslant \ C_{st}\, \mathbb{E}\Big[\Big\| \sum_{\ell=1}^{m_n u_n} \Psi_{n,\ell-1}\, \varepsilon_\ell \Big\|^4\Big] \\
&= \ O\big((m_n\, u_n\, |||B_n^{-1}|||)^2\big)
\end{aligned}$$

as n tends to infinity. This result is reached using the strict stationarity of the process, the explicit expression of $X_{n,0}^4$ in terms of A_n^ℓ, the inequality (3.6) and, finally, using (H_4) giving the equivalence between $(1-\rho(A_n))^{-2}$ and $C_{st}\, |||B_n^{-1}|||^2$. So,

$$\mathbb{E}\big[\langle \lambda,\, \xi''_{n,1} \rangle^4\big] = O(m_n^2\, u_n^2).$$

By Lyapunov's inequality,

$$\mathbb{E}\big[|\langle \lambda,\, \xi''_{n,1} \rangle|^{3+\delta}\big] \ \leqslant \ \big(\mathbb{E}\big[\langle \lambda,\, \xi''_{n,1} \rangle^4\big]\big)^{\frac{3+\delta}{4}} \ = \ O\Big((m_n\, u_n)^{\frac{3+\delta}{2}}\Big)$$

for a small $\delta > 0$. Now, combining this result with (3.25) and Hölder's inequality, for the sufficiently large values of n,

$$\begin{aligned}
\frac{t_n\, b_n}{n^{\frac{3}{2}}}\, \mathbb{E}\big[|\langle \lambda,\, \xi''_{n,1} \rangle^3\, \mathrm{e}^{\nu_n}|\big] \ &\leqslant \ \frac{t_n\, b_n}{n^{\frac{3}{2}}}\, \big(\mathbb{E}\big[|\langle \lambda,\, \xi''_{n,1} \rangle|^{3+\delta}\big]\big)^{\frac{3}{3+\delta}}\, \big(\mathbb{E}\big[\mathrm{e}^{\frac{3+\delta}{\delta}\,\nu_n}\big]\big)^{\frac{\delta}{3+\delta}} \\
&\leqslant \ C_{st}\, \frac{t_n\, b_n}{n^{\frac{3}{2}}}\, (m_n\, u_n)^{\frac{3}{2}} \ \longrightarrow \ 0 \qquad (3.27)
\end{aligned}$$

by (3.25), (3.23) and the properties in (3.22). The second-order term in (3.26) satisfies

$$\begin{aligned}
\frac{t_n}{2\,n}\, \mathbb{E}\big[\langle \lambda,\, \xi''_{n,1} \rangle^2\big] \ = \ \frac{t_n}{2\,n}\, \lambda^T\, \mathbb{V}(\xi''_{n,1})\lambda \ &= \ \frac{t_n\, u_n\, m_n}{2\,n\, |||B_n^{-1}|||^3}\, \lambda^T\, B_n^{-1}\, \mathbb{V}(\mathrm{vec}(\zeta_{n,1}))(B_n^{-1})^T \lambda \\
&= \ \frac{t_n\, u_n\, m_n}{2\,n}\, \lambda^T\, \frac{B_n^{-1}\, \Upsilon_n\, (B_n^{-1})^T}{|||B_n^{-1}|||^3}\, \lambda \\
&\longrightarrow \ \frac{1}{2}\, \langle \lambda, \Upsilon\, \lambda \rangle \qquad (3.28)
\end{aligned}$$

where we used (3.23) and the results of Lemma 3.6. The combination of (3.26), (3.27) and (3.28) together with the Gärtner-Ellis theorem (see *e.g.* Sec. 2.3 of [8]) shows that the sequence

$$\left(\frac{1}{b_n \sqrt{n}} \sum_{\ell=1}^{n} \xi_{n,\ell} \right)_{n \geqslant 1}$$

satisfies an LDP with speed (b_n^2) and rate function given by the Fenchel-Legendre transform of the above logarithmic moment generating function, *i.e.*

$$I(x) = \sup_{\lambda \in \mathbb{R}^{p^2}} \Big\{ \langle \lambda, x \rangle - \frac{1}{2}\, \langle \lambda, \Upsilon\, \lambda \rangle \Big\}.$$

Note that, due to its particular structure, Υ is only non-negative definite as soon as $p > 1$ (by way of example, its last row and column are zero). In that case (see *e.g.* Ex. 1.1.4 of [12]), the explicit expression of this quadratic rate function, strictly convex on its relative interior, is

$$I(x) = \left\{ \begin{array}{ll} \frac{1}{2}\, \langle x, \Upsilon^{\dagger}\, x \rangle & \text{for } x \in \mathrm{Im}(\Upsilon) = \mathrm{Im}(\Upsilon^{\dagger}) \\ +\infty & \text{otherwise} \end{array} \right.$$

where $\Upsilon^{\dagger}$ denotes the generalized inverse of Υ. After the truncation introduced in (3.14), the decomposition (3.9) can be rewritten as

$$\begin{aligned} \frac{\sqrt{n}\, (1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, \mathrm{vec}\bigg(\frac{1}{n} \sum_{k=1}^{n} (\Phi_{n,\,k}\, \Phi_{n,\,k}^{T} - \Gamma_n) \bigg) \ &= \ \frac{(1-\rho(A_n))^{\frac{3}{2}}\, |||B_n^{-1}|||^{\frac{3}{2}}}{b_n \sqrt{n}} \sum_{k=1}^{n} \xi_{n,\,k} \\ &\quad + \frac{\sqrt{n}\, (1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, R_n^{*} \end{aligned}$$

where, in the remainder term $R_n^{*} = B_n^{-1}\, \mathrm{vec}(\Lambda_n) - R_n$, the residual of the truncation is given in (3.19) and the main residual R_n is given in (3.11). Lemma 3.4 and Lemma 3.7 show that the first term in the right-hand is an exponentially good approximation of the left-hand side and that, as a consequence, they share the same LDP (see Def. 4.2.10 and Thm. 4.2.13 of [8]). The contraction principle (see Thm. 4.2.1 of [8]) enables to compute the rate function associated with the LDP, namely

$$I_{\Gamma}(x) = I\big(h^{-\frac{3}{2}}\, x\big) = \left\{ \begin{array}{ll} \frac{1}{2\, h^3}\, \langle x, \Upsilon^{\dagger}\, x \rangle & \text{for } x \in \mathrm{Im}(\Upsilon) = \mathrm{Im}(\Upsilon^{\dagger}) \\ +\infty & \text{otherwise} \end{array} \right. \tag{3.29}$$

where the limiting value $h > 0$ comes from (H_4). □

3.2.6. *Proof of Theorem 2.2.* Taking back the notations (2.2) and (2.3),

$$\begin{aligned} \frac{\sqrt{n}}{b_n\, (1-\rho(A_n))^{\frac{1}{2}}}\, \big(\widehat{\theta}_n^{\,\pi} - \theta_n^{\pi}\big) \ &= \ \frac{\sqrt{n}\, (S_{n-1}^{\pi})^{-1}}{b_n\, (1-\rho(A_n))^{\frac{1}{2}}} \sum_{k=1}^{n} \Phi_{n,\,k-1}\, \varepsilon_k \\ &= \ \frac{n\, |||B_n^{-1}|||\, (S_{n-1}^{\pi})^{-1}}{b_n \sqrt{n}\, |||B_n^{-1}|||^{\frac{1}{2}}\, (1-\rho(A_n))^{\frac{1}{2}}\, |||B_n^{-1}|||^{\frac{1}{2}}} \sum_{k=1}^{n} \Phi_{n,\,k-1}\, \varepsilon_k. \end{aligned}$$

Our objective is first to prove that, for all $r > 0$,

$$\lim_{n \to +\infty} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(|||n\, |||B_n^{-1}|||\, (S_{n-1}^{\pi})^{-1} - \Gamma_{\pi}^{-1}||| \geqslant r \bigg) = -\infty \tag{3.30}$$

where Γ_{π} is the invertible penalized variance (2.3), and then to establish an LDP for the sequence

$$\left(\frac{1}{b_n \sqrt{n}\, |||B_n^{-1}|||^{\frac{1}{2}}} \sum_{k=1}^{n} \Phi_{n,\,k-1}\, \varepsilon_k \right)_{n \geqslant 1} \tag{3.31}$$

so as to come to the announced result, *via* the contraction principle (Thm. 4.2.1 of [8]). On the one hand, we know from Theorem 2.1 and (3.29) that

$$\frac{1}{b_n^2} \ln \mathbb{P}\bigg(\bigg|\bigg|\bigg| \frac{S_{n-1}}{n\, |||B_n^{-1}|||} - \frac{\Gamma_n}{|||B_n^{-1}|||} \bigg|\bigg|\bigg| \geqslant r \bigg) \ = \ \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\frac{\sqrt{n}}{b_n\, |||B_n^{-1}|||^{\frac{3}{2}}} \bigg|\bigg|\bigg| \frac{S_{n-1}}{n} - \Gamma_n \bigg|\bigg|\bigg| \geqslant r_n \bigg)$$

$$\longrightarrow \ -\infty \ = \ - \lim_{\|x\| \to +\infty} I_\Gamma(x)$$

since, by (H_4) and (H_5),

$$r_n = \frac{r\sqrt{n}}{b_n \, |||B_n^{-1}|||^{\frac{1}{2}}} \ \longrightarrow \ +\infty$$

and $(1-\rho(A_n))^{\frac{3}{2}} \sim h^{\frac{3}{2}} \, |||B_n^{-1}|||^{-\frac{3}{2}}$. So,

$$\lim_{n \to +\infty} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\bigg|\bigg|\bigg| \frac{S_{n-1}^\pi}{n\,|||B_n^{-1}|||} - \Gamma_n^\pi \bigg|\bigg|\bigg| \geqslant r \bigg) = -\infty \quad \text{for} \quad \Gamma_n^\pi = \frac{\Gamma_n}{|||B_n^{-1}|||} + \pi\, I_p.$$

It is also clear that

$$\bigg\{ \bigg|\bigg|\bigg| \frac{S_{n-1}^\pi}{n\,|||B_n^{-1}|||} - \Gamma_\pi \bigg|\bigg|\bigg| \geqslant r \bigg\} \ \subset \ \bigg\{ \bigg|\bigg|\bigg| \frac{S_{n-1}^\pi}{n\,|||B_n^{-1}|||} - \Gamma_n^\pi \bigg|\bigg|\bigg| \geqslant \frac{r}{2} \bigg\} \cup \bigg\{ |||\Gamma_n^\pi - \Gamma_\pi||| \geqslant \frac{r}{2} \bigg\}$$

and (1.11) shows that the second event in the right-hand side becomes impossible when n increases. Hence, from the reasoning above,

$$\lim_{n \to +\infty} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\bigg|\bigg|\bigg| \frac{S_{n-1}^\pi}{n\,|||B_n^{-1}|||} - \Gamma_\pi \bigg|\bigg|\bigg| \geqslant r \bigg) = -\infty.$$

Now we shall use Lem. 2 of [21] to get (3.30). On the other hand, all the work to prove that the sequence (3.31) satisfies an LDP with speed (b_n^2) is done in the proof of Theorem 2.1. Indeed, *via* the truncation (3.14),

$$\begin{aligned} \frac{1}{b_n \sqrt{n}\, |||B_n^{-1}|||^{\frac{1}{2}}} \sum_{k=1}^{n} \Psi_{n,\,k-1}\, \varepsilon_k \ &= \ \frac{1}{b_n \sqrt{n}\, |||B_n^{-1}|||^{\frac{1}{2}}} \sum_{k=1}^{n} \sum_{\ell=0}^{m_n-2} A_n^\ell \, E_{k-\ell-1}\, \varepsilon_k \\ &= \ \frac{1}{b_n \sqrt{n}} \sum_{k=1}^{n} Z_{n,\,k} \end{aligned}$$

where the process $(Z_{n,\,k})_k$ forms a strictly stationary and m_n–dependent sequence. However, apart from the renormalization, this is precisely the first column of the first term of (3.15). Thus, the calculations are similar and we find, like in Lemma 3.6,

$$\mathbb{V}(Z_{n,\,1}) = \frac{\sigma^2\, \Gamma_{n,\,m_n}}{|||B_n^{-1}|||}.$$

In that case, from the convergence (3.17) and the previous proof, the rate function associated with the LDP is given by

$$J(x) = \sup_{\lambda \in \mathbb{R}^p} \Big\{ \langle \lambda, x \rangle - \frac{\sigma^2}{2} \, \langle \lambda, \Gamma\, \lambda \rangle \Big\} = \begin{cases} \frac{1}{2\,\sigma^2} \, \langle x, \Gamma^\dagger \, x \rangle & \text{for } x \in \mathrm{Im}(\Gamma) = \mathrm{Im}(\Gamma^\dagger) \\ +\infty & \text{otherwise} \end{cases}$$

where $\Gamma^\dagger$ denotes the generalized inverse of Γ. The exponential negligibility of the remainder of the truncation is obtained following the lines of Lemma 3.7. The contraction principle enables to compute the rate function associated with the LDP, namely

$$I_\theta(x) = J\big(\Gamma_\pi \sqrt{h}\, x\big) = \begin{cases} \frac{h}{2\,\sigma^2} \, \langle x, \Gamma_\pi\, \Gamma^\dagger\, \Gamma_\pi \, x \rangle & \text{for } x \in \mathrm{Im}(\Gamma_\pi^{-1}\, \Gamma) \\ +\infty & \text{otherwise} \end{cases} \tag{3.32}$$

where the exponential convergence (3.30) has been combined to the LDP established on the sequence (3.31), the limiting value $h > 0$ coming from (H_4). $\square$

References

[1] Bercu, B. On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases. *Bernoulli. 7* (2001), 299–316.

[2] Bercu, B., Gamboa, F., and Rouault, A. Large deviations for quadratic forms of stationary Gaussian processes. *Stoch. Proc. Appl. 71* (1997), 75–90.

[3] Bitseki Penda, V., Djellout, H., and Proïa, F. Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *ESAIM Probab. Stat. 18* (2014), 308–331.

[4] Brockwell, P. J., and Davis, R. A. *Time series: Theory and Methods (Second Edition).* Springer Series in Statistics. Springer, New York, 1991.

[5] Buchmann, B., and Chan, N. H. Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Proc. Appl. 123* (2013), 952–985.

[6] Chan, N. H., and Wei, C. Z. Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Stat. 15* (1987), 1050–1063.

[7] Chan, N. H., and Wei, C. Z. Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist. 16* (1988), 367–401.

[8] Dembo, A., and Zeitouni, O. *Large Deviations Techniques and Applications (Second Edition)*, vol. 38 of *Applications of Mathematics.* Springer, 1998.

[9] Djellout, H., Guillin, A., and Wu, L. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. I. H. Poincaré. 42* (2006), 393–416.

[10] Donsker, M. D., and Varadhan, S. R. S. Large deviations for stationary Gaussian processes. *Comm. Math. Phys. 97* (1985), 187–210.

[11] Duflo, M. *Random iterative models.* Applications of Mathematics (vol. 34), New York. Springer-Verlag, Berlin, 1997.

[12] Hiriart-Urruty, J. B., and Lemaréchal, C. *Fundamentals of Convex Analysis.* Grundlehren Text Editions. Springer, 2012.

[13] Horn, R. A., and Johnson, C. R. *Matrix Analysis (Second Edition).* Cambridge University Press, Cambridge, New-York, 2012.

[14] Jiang, T., and Wei, M. On solutions of the matrix equations $X - AXB = C$ and $X - A\bar{X}B = C$. *Linear Algebra Appl. 367* (2003), 225–233.

[15] Mas, A., and Menneteau, L. Large and moderate deviations for infinite-dimensional autoregressive processes. *J. Multivariate Anal. 87* (2003), 241–260.

[16] Miao, Y., and Shen, S. Moderate deviation principle for autoregressive processes. *J. Multivariate Anal. 100* (2009), 1952–1961.

[17] Miao, Y., Wang, Y., and Yang, G. Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat. 42* (2015), 234–255.

[18] Miao, Y., and Yang, G. A moderate deviation principle for m–dependent random variables with unbounded m. *Acta Appl. Math. 104* (2008), 191–199.

[19] Phillips, P. C. B., and Lee, J. H. Limit theory for vars with mixed roots near unity. *Economet. Rev. 34* (2015), 1034–1055.

[20] Phillips, P. C. B., and Magdalinos, T. Limit theory for moderate deviations from a unit root. *J. Econometrics. 136* (2007), 115–130.

[21] Worms, J. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab. 4* (1999), 1–28.

[22] Wu, W. B., and Zhao, Z. Moderate deviations for stationary processes. *Stat. Sinica. 18* (2008), 769–782.

Laboratoire angevin de recherche en mathématiques, LAREMA, UMR 6093, CNRS, UNIV Angers, SFR MathSTIC, 2 Bd Lavoisier, 49045 Angers Cedex 01, France.

E-mail address: `frederic.proia@univ-angers.fr`